Application No. 10/049,650   Page 2
Response

Remarks

Claims 1-11 are pending in the application. Claims 1-11 were rejected. Claim 1 is the independent claim. Reconsideration of the rejection in view of the following remarks is respectfully requested.

The examiner objected to claim 5 under 37 CFR 1.75(c) as being in improper form because a multiple dependent claim should refer to other claims in the alternative only. It is respectfully pointed out that claim 5 is not a multiple dependent claim. Claim 5 makes reference to sub-elements a1-3) and b1-3), which are originally recited in claims 1 and 4, respectively. However, claim 5 depends directly from claim 4, which in turn depends from claim 1. Therefore, each of these sub-elements is recited in claim 5 by incorporation, and claim 5 is not a multiple dependent claim. The objection, therefore, should be withdrawn.

The examiner rejected claims 1-11 under 35 USC §102(b) as being anticipated by Mehrotra et al.

Independent claim 1 recites a method for training a neural network in order to optimize the structure of the neural network. The neural network includes an input layer having a plurality of input neurons that receive the input data, at least one intermediate layer having a plurality of intermediate neurons, an output layer having a plurality of output neurons that provide output signals, and a multiplicity of synapses, wherein each synapse interconnects two neurons of different layers, defining a sending direction from the input layer toward the output layer. The training method includes identifying and eliminating synapses that have no significant influence on the curve of the risk function. First and second sending neurons are selected that are connected to the same receiving

BEST AVAILABLE COPY

Application No. 10/049,650   Page 3
Response

neuron by respective first and second synapses. It is assumed that there is a correlation of

response signals from the first and second sending neurons to the same receiving neuron.

The first synapse is interrupted and a weight of the second synapse is adapted in its place.

The output signals of the changed neural network are compared with the output signals of

the unchanged neural network. If the comparison result does not exceed a predetermined

level, the first synapse is eliminated, thereby simplifying the structure of the neural

network.

In contrast, Mehrotra et al. disclose a conventional neural network pruning

process. Mehrotra et al. first disclose three approaches to building an optimally-sized

network, one of which is to prune the network by elimination of nodes and connections

that can be considered unimportant (page 116, lines 6-9). In Section 4.2.1, Mehrotra et

al. describe four pruning procedures in more detail, as follows.

1. Elimination of connections associated with weights of small magnitude, and

elimination of nodes having associated connections with small magnitude weights.

2. Elimination of connections that do not significantly affect network outputs or

error. The test for these connections consists of replacing the connection weight with

zero or some other weight and detecting the change in network output.

3. Elimination of input nodes if the resulting change in network output is

negligible.

4. Determination of which weights can be eliminated by examining second

derivatives of the error function contained in the Hessian matrix.

All of the disclosed procedures eliminate nodes and/or connections from a network,

with the goal of optimizing network size. However, in determining which nodes and

Application No. 10/049,650   Page 4
Response

connections to eliminate, Mehrotra et al. rely on an assessment of the weight of the

element being considered, empirical results of elimination, or examination of an error

function. None of these procedures includes selecting first and second sending neurons

that are connected to the same receiving neuron by respective first and second synapses,

assuming a correlation of response signals from said first and second sending neurons to

the same receiving neuron, interrupting the first synapse and adapting in its place a

weight of the second synapse, comparing the output signals of the neural network

changed in accordance with the previous action with the output signals of the unchanged

neural network, and if the comparison result does not exceed a predetermined level,

eliminating the first synapse, thereby simplifying the structure of the neural network. All

of these actions are required by claim 1, and Mehrotra et al. do not disclose any of them.

In the detailed rejection, the examiner asserts that Mehrotra et al. disclose

selecting first and second sending neurons that are connected to the same receiving

neuron by respective first and second synapses on page 19, in Fig. 1.14, and on page 116,

at lines 31 and 32. Fig. 1.14 shows the layout of an acyclic network. The passage on

page 116 merely states that nodes having associated connections with small magnitude

weights can be eliminated. The passage does not disclose selecting first and second

sending neurons connected to the same receiving neuron, nor does the disclosed pruning

process require more than one sending neuron connected to a receiving neuron for

pruning to take place. This claim element is not disclosed by Mehrotra et al.

The examiner further asserted that Mehrotra et al. disclose assuming a correlation

of response signals from said first and second sending neurons to the same receiving

neuron on page 19, in Fig. 1.14, and on page 46, in Fig. 2.5. Fig. 2.5 shows a generic

Application No. 10/049,650   Page 5
Response

perceptron for *n*-dimensional input space, implementing the hyperplane in which the sum of the products of all input nodes and their respective weights is equal to zero, which produces a step function. This figure in particular is used to illustrate a hyperplane that can separate samples of different classes for spaces of higher numbers of input dimensions, for purposes of training a network, not for pruning a network. To the extent that the disclosed relation might suggest a correlation of response signals from first and second sending neurons, it is not applicable to the pruning procedure described on pages 116-118. This claim element is not disclosed by Mehrotra et al.

The examiner asserted that Mehrotra et al. disclose interrupting the first synapse and adapting in its place a weight of the second synapse on page 19, in Fig. 1.14, and on page 117, in Fig. 4.6. The examiner noted that the weight of the second synapse is the result of the removal of the first synapse. Fig. 4.6 discloses a generic network pruning algorithm. According to this algorithm, a node or connection is found, the removal of which does not penalize performance beyond desirable tolerance levels. The node or connection is deleted. The algorithm does not disclose interrupting a first synapse and adapting in its place a weight of the second synapse. Actually, the algorithm does not require first and second synapses at all, and only describes analysis of a single node or connection to make a pruning determination. Further, there is no support for the examiner's note that removal of a first synapse results in replacement by the weight of a second synapse. This claim element is not disclosed by Mehrotra et al.

The examiner also asserted that Mehrotra et al. disclose comparing the output signals of the neural network changed in accordance with the previous action with the output signals of the unchanged neural network on page 19, in Fig. 1.14, on page 116, at

Application No. 10/049,650   Page 6
Response

lines 33 and 34, and on page 17, in Fig. 4.6. The passage on page 116 merely states that

connections having an existence that does not significantly affect network output or error

may be pruned. That is, the cited passage discloses a goal of pruning, but does not

disclose the process of eliminating synapses recited in claim 1. As previously noted, Fig.

4.6 discloses an algorithm that includes elimination of a node or connection such that the

elimination does not penalize performance beyond a desirable tolerance level. Again,

this algorithm represents a goal of a pruning procedure, but does not disclose the process

of eliminating synapses recited in claim 1. This claim element is not disclosed by

Mehrotra et al.

The examiner also asserted that Mehrotra et al. disclose eliminating the first

synapse if the comparison result does not exceed a predetermined level, thereby

simplifying the structure of the neural network on page 19, in Fig. 1.14 and on page 118

at lines 1-5. The passage on page 118 is a small portion of a description of the Lecun,

Denker, and Solla Optimal Brain Damage method of pruning a network, and therefore is

unrelated to the other passages cited by the examiner in the Mehrotra et al. reference.

That is, this cited action is not part of the same process described in the earlier passages,

and therefore, taken out of context, should not be applied against a single element in

claim 1 when other processes have been applied against other elements of the claim, in an

attempt to demonstrate anticipation of the claimed invention. In any case, the passage on

page 118 merely states that a connection may be eliminated if the result of the

elimination results in an error that is no greater than that expressed by the formula on line

3. That is, a threshold tolerance is expressed for pruning network connections. Again,

this is just a goal of the pruning process, albeit this time with a specifically-stated

Application No. 10/049,650   Page 7
Response

tolerance level. However, this tolerance level is related to an error resulting from elimination of a connection, and not related to an error resulting from adapting the weight of a second synapse in place of that of a first synapse, as recited in claim 1. This claim element is not disclosed by Mehrotra et al.

In summary, Mehrotra et al. disclose that pruning of networks is advantageous, and that it is preferable to establish a tolerance level for errors resulting from the pruning procedure. However, Mehrotra et al. do not disclose the particular synapse identification and elimination method recited in claim 1. Rather, Mehrotra et al. disclose elimination of connections and nodes based on their weights alone, and analysis of the errors caused by these singular eliminations. Mehrotra et al. do not perform the selection, interruption, and comparison actions recited in claim 1. The process recited in claim 1 is limited to receiving neurons having at least two connected sending neurons, whereas Mehrotra et al. allow pruning to take place at any node, because comparison between two synapses is not necessary or performed according to Mehrotra et al.

In view of the discussion set forth above, it is submitted that Mehrotra et al. do not anticipate the invention recited in claim 1. Claims 2-11 depend from claim 1, and therefore also are not anticipated by Mehrotra et al. The rejection of claims 1-11, therefore, should be withdrawn.

The examiner addresses comments particularly to the remarks submitted in response to the previous Office action, which are discussed below.

The examiner pointed out that neural networks are massively parallel devices, and that it is common for at least two sending neurons to be connected to a third, receiving neuron, and that correlation among sending neurons is shown by Mehrotra et al. This is

Application No. 10/049,650   Page 8
Response

not disputed. However, the claimed invention requires <u>selection</u> of two particular

sending neurons having correlated response signals as part of the synapse identification

and elimination process. Mehrotra et al. acknowledge that such neurons exist, but do not

perform the claimed selection action when pruning the network. That is, although

Mehrotra et al. acknowledge pluralities of correlated sending neurons, that characteristic

is irrelevant to the disclosed pruning process, whereas it is essential to the claimed

process.

On page 8 of the action, the examiner notes that, according to Mehrotra et al., if

by chance a connection is eliminated and the receiving neuron only receives two

connections, then necessarily the only remaining element to the receiving neuron is the

other perceptron, which is influenced by the weight of the second perceptron. Accepting

this proposition for the sake of discussion, the elimination of one of only a pair of

sending connections would only happen by chance according to the Mehrotra et al.

process, whereas the claimed process explicitly includes a <u>selection</u> of sending neuron

pairs. No such selection takes place according to Mehrotra et al. Rather, connections and

nodes are selected according to their weights, so that connections associated with weights

of small magnitude are eliminated. Selection of sending neuron pairs is not disclosed by

Mehrotra et al.

At the bottom of page 8, the examiner concluded that Mehrotra et al., at paragraph

4.2.1 and Fig. 4.6, disclose elimination of nodes having a weight below a threshold. This

is the same conclusion in the applicants' argument, which the examiner quoted above his

response. That is, Mehrotra et al. select connections for elimination based on connection

<u>weight</u>, not on analysis based on selected sending neuron pairs. The claimed invention

Application No. 10/049,650   Page 9
Response

does not select connections based on weight; likewise, the existence of neuron pairs is irrelevant to Mehrotra et al., although it is required by claim 1.

On page 9, the examiner asserts that the Mehrotra et al. network is made of a plurality of neurons and therefore there is a plurality of pairs of neurons and such pairs will be selected for weight adaptation. However, this selection is not disclosed by Mehrotra et al. In fact, Mehrotra et al. never disclose selection of more than one connection at a time. Mehrotra et al. disclose selection of individual connections based on weight, and do not disclose selection of pairs for any reason under any circumstances. It is clear from the Mehrotra et al. disclosure that pairs exist, but they are not selected. This remark also applies to the examiner's comments on page 10.

It is acknowledged that the examiner has full latitude to interpret the claims in their broadest reasonable sense. However, it is respectfully submitted that no reasonable interpretation of "selecting first and second sending neurons that are connected to the same receiving neuron by respective first and second synapses" could be so broad as to cover a process in which only individual neurons are selected according to weight, even though each might coincidentally be one of a first and second neuron.

Application No. 10/049,650   Page 10
Response

Based on the foregoing, it is submitted that all objections and rejections have been

overcome. It is therefore requested that the claims be allowed, and the case passed to

issue.

Respectfully submitted,

March 10, 2005   Thomas M. Champagne
Date   Registration No. 36,478
                                                 IP STRATEGIES
                                                 12 ½ Wall Street
                                                 Suite I
                                                 Asheville, North Carolina 28805
                                                 828.253.8600
TMC:hlp   828.253.8620 fax